John von Neumann Institute for Computing

Protein-Ligand Docking with Two-Scale Receptor Dynamics and QM/MM Potential

J. A. Kalinowski, B. Lesyng

published in

From Computational Biophysics to Systems Biology (CBSB08),
Proceedings of the NIC Workshop 2008,
Ulrich H. E. Hansmann, Jan H. Meinke, Sandipan Mohanty,
Walter Nadler, Olav Zimmermann (Editors),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **40**, ISBN 978-3-9810843-6-8, pp. 117-120, 2008.

http://www.fz-juelich.de/nic-series/volume40

Protein-Ligand Docking with Two-Scale Receptor Dynamics and QM/MM Potential

Jarosław A. Kalinowski and Bogdan Lesyng

CoE BioExploratorium and Department of Biophysics, University of Warsaw,
Zwirki i Wigury 93, 02-089 Warsaw, Poland
E-mail: jak@biogeo.uw.edu.pl, lesyng@icm.edu.pl

A new docking model PhDock is presented. The docking process is correlated with global deformations (normal modes) of a receptor and with reorientations of its side chains. Deformations are derived from an elastic network model. Microscale motions of individual residues are generated using a statistical potential derived from a library of conformers. A basin-hopping Monte Carlo dynamics of a ligand is carried out using the SCC-DFTB energy plus protein - ligand and Lennard-Jones interactions. The electrostatic contribution is computed using SCC-DFTB atomic charges (Mulliken and CM3) in the presence of the receptor field, which accounts for the ligand polarization effects.

1 Introduction

Docking methods of small molecules to protein targets have been extensively developed in the past few decades. Due to constant increase of available computational power, some procedures that were previously out of reach become presently possible. We present a flexible-receptor and flexible-target model that is based on extensive two-scale probing of receptor configurations, and the ligand dynamics which is based on a QM/MM potential. The slow-motion, global receptor deformations, are accounted for by considering first few target normal modes[1] derived from an elastic network model[2]. In turn, faster reorientations of amino acid side chains allow achieving more energetically favorable interactions by the ligand. This reorientations are modelled by switching discreet side chain conformations, taken from a library[3], according to a Monte Carlo dynamics using library provided propensities as a stochastic potential. Mezoscopic and microscopic motions are decoupled, except for the checking of sterical conflicts. In the presence of the receptor with its own dynamics, a basin-hopping Monte Carlo ligand dynamics is carried out with the potential that combines ligand deformation energy computed using a fast SCC-DFTB QM method, together with protein-ligand coupling given by Coulomb and Lennard-Jones interactions with the receptor atoms. The electrostatic contribution is computed using ligand CM3/SCC-DFTB atomic charges[4] in the presence of the receptor, which accounts for electron polarization effects.

2 Potential Energy Function

The potential used for the ligand minimization and dynamics is defined as follows:

$$E = E^{LJ} + E^{ES}[q^{\mathrm{CM3+P}}; \epsilon] + E^{QM} \tag{1}$$

where E^{LJ} is the Lennard-Jones term computed using Amber94 parameters for the receptor, and a fixed set of LJ parameters for the ligand atoms. The remaining terms represent

the Coulomb interaction energy with the CM3 charges and polarization corrections to the ligand charges, as well as the ligand quantum mechanical energy, which is computed using the SCC-DFTB method, which has an algebraic structure of a Tight-Binding method with the energy decomposition into the "band structure" and "repulsion":

$$E = \sum_{i}^{\text{occ.}} \epsilon_i + \frac{1}{2} \sum_{\text{A}\neq\text{B}} U_{\text{AB}}(|R_{\text{A}} - R_{\text{B}}|). \tag{2}$$

The Hamiltonian, however, has an extra term $H^1_{\mu\nu}$ which includes interactions of intramolecular partial charges and interaction with external electrostatic field:

$$H^{\text{SCC-DFTB}}_{\mu\nu} := H^{\text{DFTB}}_{\mu\nu} + H^1_{\mu\nu} \qquad H^1_{\mu\nu} = S_{\mu\nu}(\Gamma_{\text{A}} + \Gamma^{\text{ext.}}_{\text{A}}), \;\; \nu \in \text{A} \tag{3}$$

where Γ_{A} and $\Gamma^{\text{ext.}}_{\text{A}}$ are corresponding potential shifts on atom A, and $S_{\mu\nu}$ is the overlap matrix. Atomic charges on the receptor are included in ligand SCC-DFTB calculations *via* $\Gamma^{\text{ext.}}_{\text{A}}$ shifts:

$$\Gamma^{\text{ext.}}_{\text{A}} = \sum_{\text{n}} \frac{1}{\epsilon} \frac{Q_{\text{n}}}{|R_{\text{A}} - R_{\text{n}}|} \tag{4}$$

where Q_{n} are the receptor charges, and ϵ scales the strength of the interaction potential.

2.1 CM3/SCC-DFTB Charges

The CM3 procedure computes a correction to the Mulliken charges after the SCC cycle, which reduces systematic errors of individual bond dipoles:

$$q_k = q^0_k + \sum_{k'\neq k} T_{kk'}(B_{kk'}) \tag{5}$$

where q_k is the CM3 charge on an atom k, q^0_k is the original Mulliken charge, $B_{kk'}$ is the Mayer bond order and $T_{kk'}$ is a function of the bond orders which determines the amount of the charge to be transferred from an atom k' to an atom k:

$$T_{kk'} = D_{Z_k Z_{k'}} B_{kk'} + C_{Z_k Z_{k'}} (B_{kk'})^2. \tag{6}$$

The C and D coefficients were determined by the parameterization procedure.

3 Sampling Scheme

3.1 Side Chain Conformer Libraries

We use libraries constructed and published by Shetty et al.[3]. These side chain conformer libraries were extracted from high-quality protein structures, they maintain crystallographic bond lengths and angles, in contrast to traditional rotamer libraries defined in terms of χ angles under the assumption of idealized covalent geometry. The libraries provide also backbone ϕ, ψ dependent propensities for each 40° dihedral angle bin, computed from the conformational populations.

The conventional Metropolis algorithm[5] is formulated as a rule which states that transition from a state A to a state B is accepted with the probability $\exp(-(E_B - E_A)/kT)$. This is, however, equal to the ratio P_A/P_B of the probabilities of states A and B. In this approach the probabilities are taken from the normalized propensities included in the SCL libraries.

Destination conformations are drawn with equal probabilities from the whole ensemble of conformations allowed by the ϕ, ψ angles. Frequency of side chain orientation changes for a given residue is determined by its distance to the ligand. Currently a dumping function $e^{-(\frac{r}{r_0})^2}$ is used with r_0 in range of 0.7...1.2 nm.

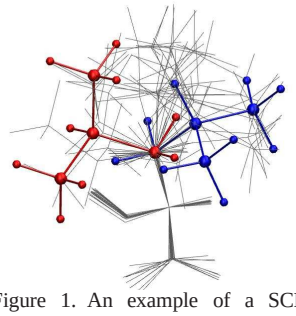

Figure 1. An example of a SCL step. Two randomly selected conformations are shown using the CPK model. Every second conformation from the SCL0.5 library is drawn using thin lines.

3.2 Normal Modes

Normal modes are computed using MMTK library[6] written by Konrad Hinsen, in the C_α representation of the protein. The applied potential is defined as follows:

$$V(\vec{r}_1 \dots \vec{r}_N) = \sum_{i<j} f(r_{ij}^0)(r_{ij} - r_{ij}^0)^2$$

where $r_{ij} = |\vec{r}_i - \vec{r}_j|$, and the pair force constant $f(r_{ij}^0)$ is given by the expression:[2]

$$k(r_{ij}^0) = \begin{cases} 8.6 \times 10^5 r_{ij}^0 - 2.39 \times 10^5 & r_{ij}^0 < 0.4\text{nm} \\ 128(r_{ij}^0)^{-6} & r_{ij}^0 \geq 0.4\text{nm} \end{cases}$$

using nm and kJ/mol units, and depends on residue distance in the reference conformation. Diagonalization of the second derivatives matrix using a fourier basis gives the modes $m_{\mu i}$ and the corresponding frequencies f_μ. The resulting atomic displacement vectors for selected modes (default: the lowest four nonzero modes) are copied from the C_α atoms to all remaining atoms in each residue.

Let the nm-state be defined as $d = (d_1 \dots d_m) \in \mathbb{R}^m$, where m is the number of selected modes and d_μ can be interpreted as a measure of deformation along the μ-th selected mode. The receptor atom i displacement is given by:

$$\vec{r}_i = \vec{r}_i^{\,0} + \sum_\mu d_\mu m_{\mu i}$$

The Monte Carlo dynamics using the Metropolis algorithm[5] is carried out with the quadratic potential $V(d) = \sum f_\mu d_\mu^2$. The temperature does not have any well defined physical meaning and is treated as a parameter of the model.

3.3 Algorithm

The algorithm consists of a sequence of steps that can be summarized as follows:

draw a step-type with probabilities: *nm:scl:rb:ligmin* (typically: 1:4:64:1 or 2:4:32:1)
if step-type is *nm* **then**
  $d^{\text{new}} = d + \delta$ where δ is a random vector in $\mathbb{R}^m$
  if step valid and accepted by `Metropolis`(E^{new}, E) **then**
    apply deformation resulting from δ to the receptor; find new basin minimum
  end if
else if step-type is *scl* **then**
  draw residue using probabilities given by the dumping function $\exp(-r^2/r_0^2)$
  draw new conformation from conformations allowed by residue ϕ, ψ angles
  if step valid and accepted with probability P_A/P_B **then**

change side chain orientation; find new basin minimum
end if
else if step-type is *rb* **then**
draw transformation being superposition of random, rigid rotation and translation
find new ligand configuration and its basin minimum $E_{\text{basin}}^{\text{new}}$
if step valid and accepted with probability $\texttt{Metropolis}(E_{\text{basin}}^{\text{new}}, E_{\text{basin}})$ **then**
accept new ligand position and new basin minimum
end if
else if step-type is *ligmin* **then**
perform fully flexible ligand minimization
end if

It should be noted the energy minimization, as a part of the basin-hopping scheme, is unrestricted allowing flexible deformations.

4 Results and Conclusion

The novel, promising multiscale QM/MM flexible docking method (PhDock) was formulated, implemented, and is being tested using a number of biomolecular model systems. These, in particular, include a short helix with several water molecules and three protein-ligand complexes with PDB codes: 1ctt, 1ivd and 2qwe. The applied stochastic sampling procedure located the correct binding modes for the first two complexes. In the last, difficult case with a deeply buried ligand, the sampling procedure located a few alternative binding poses on the protein surface. Longer sampling procedures along with refinement of selected control parameters of the model are being carried out.

Acknowledgments

These studies are supported by the N207 022 31/1172 grant and by the Centre of Excellence BioExploratorium.

References

1. M. Zacharias *Rapid Protein-Ligand Docking Using Soft Modes From Molecular Dynamics Simulations to Account for Protein Deformability: Binding of FK506 to FKBP* Proteins **54**, 759–767, 2004.
2. K. Hinsen, A.J. Petrescu, S. Dellerue, M.C. Bellissent-Funel, G. R. Kneller *Harmonicity in slow protein dynamics* Proteins **261**, 25-37, 2000.
3. R. P. Shetty, P. I.W.de Bakker, M.A. DePristo, T.L. Blundell *Advantages of fine-grained side chain conformer libraries* Protein Engineering **16**, 963–969, 2003.
4. J.A. Kalinowski, B. Lesyng, J.D. Thompson, C.J. Cramer, D.G. Truhlar *Class IV Charge Model for the Self-Consistent Charge Density-Functional Tight-Binding Method* J. Phys. Chem. A **108**, 2545–2549, 2004.
5. Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, and Augusta H. Teller *Equation of State Calculations by Fast Computing Machines* J. Chem. Phys. **21**, 1087–1092, 1953.
6. K. Hinsen *The Molecular Modeling Toolkit: A New Approach to Molecular Simulations* J. Comp. Chem. **21**, 79–85, 2000.